-Title: When in Rome: I do not understand / know this expression

-Abstract: interplay between gene expression and RNA processing. Change into interplay between transcription and RNA processing

-Abstract: atomary: what does it mean in the context. Strange to be in the same sentence as molecular

-Abstract: first comprehensive collection of alternative exons and splice sites controlled by DNA variants. Shouldn’t this be common DNA variants? There are already many examples of splice site disrupting disease causing mutations archived in different databases.

-It is not clear from the abstract whether the main focus is on effects between populations or within populations. The evolutionary aspects of the paper are described in a bit of a shallow way. Either expand or reduce. It is a bit of a mixed message now. The abstract could also describe what we are going to use the knowledge for.

-The introduction better starts with the biological problem / question. By starting with GEUVADIS, you give the opinion that we have a dataset and we sought a biological problem for it.

-Intro: do not understand this sentence: a role in propagating… reference 14

-Figure 1a: that was also one of the suppl figures of the main paper, correct. I do not think this is acceptable.

-The ageing component is problematic since it is confounding with population effects. It is likely to affect alternative splicing as well. Can you use differences between cell lines from the same population to estimate / correct the extent of the problem? In general, I would put less emphasis on this.

-You use the term transcription rates but you have not measured the actual transcription rate

-However, observations of ubiquitous major transcripts…. Do not understand this sentence. Rephrase.

-I have never heard about Bhattacharyya distance. Explain?

-Dispersion is a relatively unknown term to the general public. When the manuscript is going to be sent to Nat Struct Biol I would describe or rephrase

-Figure 2C seems to contradict figure 1C of the main paper. Probably because the major transcript is considered here? It is a dangerous to present apparent conflicting results (although they may not really be contradictory)

-Figure 2D: the percentage surface proteins in the category of genes with high dispersion / or DE is extremely high. I can hardly believe these percentages

-I would be interested in genetically determined splicing differences that are in ESE and other splice signals but this has not been addressed in the paper. Findings here can lead to a better definition of these vaguely defined motifs.

-Figure 3: I wonder if you do not underestimate the frequency of activating variants, since you are looking at annotated exons only

-I miss how the variants affecting splicing and the new exons are made available.

-I find it a bit difficult that you talk about at novel introns, while most will talk about novel exons

-I miss a connection between Figure 4AB and CD.

-The coverage of RNA processing events other than alternative splicing is very shallow. Alternative first exon usage is not discussed at all.

-The number of mapped PCS is really low. Admittedly, RNA-seq is not good for polyA site detection, but to use only reads with polyA stretch is not the way to go. The way how we dealt with this problem is to use experimentally validated polyA sites (see UCSC track) and calculate coverage before and after this signal. This worked quite well and we could experimentally validate changes in polyA sites.

-I wonder why the tweedie package was chosen for DE analysis. For the per gene case with such a sample size, it is expected to be well modeled by a log-normal distribution.